Financial Econometrics
Shangai Jiao Tong University Summer School 2025-2026
Problem Set - Lecture 1

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition, Global Edition)

Review of Probability

Question 1. The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working-age population of South Africa.

	Unemployed ($Y = 0$)	Employed ($Y = 1$)	Total
Non-college grads ($X = 0$)	0.078	0.673	0.751
College grads ($X = 1$)	0.042	0.207	0.249
Total	0.12	0.88	1.000

(a) Compute $E(Y)$.

(b) The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by $1 - E(Y)$.

(c) Calculate $E(Y|X = 1)$ and $E(Y|X = 0)$.

(d) Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.

(e) A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non-college graduate?

(f) Are educational achievement and employment status independent? Explain.

Question 2. X and Y are discrete random variables with the following joint distribution:

	Value of Y				
Value of X	2	4	6	8	10
3	0.04	0.09	0.03	0.12	0.01
6	0.10	0.06	0.15	0.03	0.02
9	0.13	0.11	0.04	0.06	0.01

That is, $\Pr(X = 3, Y = 2) = 0.04$, and so forth.

(a) Calculate the probability distribution, mean, and variance of Y.

(b) Calculate the probability distribution, mean, and variance of Y given $X = 6$.

(c) Calculate the covariance and correlation between X and Y.

Review of Statistics

Question 3. In a poll of 500 likely voters, 270 responded that they would vote for the candidate from the democratic party, while 230 responded that they would vote for the candidate from the republican party. Let p denote the fraction of all likely voters who preferred the democratic candidate at the time of the poll, and let $\hat{p}$ be the fraction of survey respondents who preferred the democratic candidate.

(a) Use the poll results to estimate p.

(b) Use the estimator of the variance of $\hat{p}$, $\hat{p}(1 - \hat{p})/n$, to calculate the standard error of your estimator.

(c) What is the p-value for the test of $H_0 : p = 0.5$ vs. $H_1 : p \neq 0.5$?

(d) What is the p-value for the test of $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$?

(e) Why do the results from (c) and (d) differ?

(f) Did the poll contain statistically significant evidence that the democratic candidate was ahead of the republican candidate at the time of the poll? Explain.

Question 4. A survey of 1000 registered voters is conducted, and the voters are asked to choose between candidate A and candidate B. Let p denote the fraction of voters in the population who prefer candidate A, and let $\hat{p}$ denote the fraction of voters in the sample who prefer candidate A.

(a) You are interested in the competing hypotheses $H_0 : p = 0.4$ vs. $H_1 : p \neq 0.4$. Suppose that you decide to reject H_0 if $|\hat{p} - 0.4| > 0.01$.

  (i) What is the size of this test?

  (ii) Compute the power of this test if $p = 0.45$.

(b) In the survey, $\hat{p} = 0.44$.

  (i) Test $H_0 : p = 0.4$ vs. $H_1 : p \neq 0.4$ using a 10% significance level.

  (ii) Test $H_0 : p = 0.4$ vs. $H_1 : p < 0.4$ using a 10% significance level.

  (iii) Construct a 90% confidence interval for p.

  (iv) Construct a 99% confidence interval for p.

  (v) Construct a 60% confidence interval for p.

(c) Suppose that the survey is carried out 30 times, using independently selected voters in each survey. For each of these 30 surveys, a 90% confidence interval for p is constructed.

  (i) What is the probability that the true value of p is contained in all 30 of these confidence intervals?

  (ii) How many of these confidence intervals do you expect to contain the true value of p?

(d) In survey jargon, the "margin of error" is $1.96 \times SE(\hat{p})$. Suppose you want to design a survey that has a margin of error of at most 0.5%. How large should n be if the survey uses simple random sampling?

Question 5. Data on fifth-grade test scores (reading and mathematics) for 400 school districts in Brussels yield average score $\bar{Y} = 712.1$ and standard deviation $s_Y = 23.2$.

(a) Construct a 90% confidence interval for the mean test score in the population.

(b) When the districts were divided into districts with small classes (< 20 students per teacher) and large classes (≥ 20 students per teacher), the following results were found:

Class Size	Average Score ($\bar{Y}$)	Std. Dev. (s_Y)	n
Small	721.8	24.4	150
Large	710.9	20.6	250

Is there statistically significant evidence that the districts with smaller classes have higher average test scores? Explain.

Question 6. Suppose $Y_i \sim$ i.i.d. $N(\mu_Y, \sigma_Y^2)$ for $i = 1, \ldots, n$. With σ_Y known, the t-statistic for testing $H_0 : \mu_Y = 0$ vs. $H_1 : \mu_Y > 0$ is $t = (\bar{Y} - 0)/SE(\bar{Y})$, where $SE(\bar{Y}) = \sigma_Y/\sqrt{n}$. Suppose $\sigma_Y = 10$ and $n = 100$, so that $SE(\bar{Y}) = 1$. Using a test with a size of 5%, the null hypothesis is rejected if $t > 1.64$.

(a) Suppose $\mu_Y = 0$, so the null hypothesis is true. What is the probability that the null hypothesis is rejected?

(b) Suppose $\mu_Y = 2$, so the alternative hypothesis is true. What is the probability that the null hypothesis is rejected?

(c) Suppose that in 90% of cases the data are drawn from a population where the null is true ($\mu_Y = 0$) and in 10% of cases the data come from a population where the alternative is true and $\mu_Y = 2$. Your data came from either the first or the second population, but you don't know which.

  (i) You compute the t-statistic. What is the probability that $t > 1.64$—that is, that you reject the null hypothesis?

  (ii) Suppose you reject the null hypothesis; that is, $t > 1.64$. What is the probability that the sample data were drawn from the $\mu_Y = 0$ population?

(d) It is hard to discover a new effective drug. Suppose 90% of new drugs are ineffective and only 10% are effective. Let Y denote the drop in the level of a specific blood toxin for a patient taking a new drug. If the drug is ineffective, $\mu_Y = 0$ and $\sigma_Y = 10$; if the drug is effective, $\mu_Y = 2$ and $\sigma_Y = 10$.

  (i) A new drug is tested on a random sample of $n = 100$ patients. What is the probability that the drug is ineffective given that $t > 1.64$ (i.e., the false positive rate)?

  (ii) Suppose the one-sided test uses the 0.5% significance level. What is the false positive rate?